



FACULTAD DE CIENCIAS BIOLÓGICAS

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

PROGRAMA DE DOCTORADO EN CIENCIAS BIOLÓGICAS MENCION ECOLOGÍA

I. ANTECEDENTES GENERALES

Nombre del Curso	: ANÁLISIS Y MANIPULACIÓN DE DATOS EN R
Sigla	: BIO4022
Carácter	: Optativo
Créditos	: 5
Profesor Encargado	: Dr. Derek Corcoran
Fechas inicio y término	: 07 de Agosto al 30 de Agosto
Días	: Martes - Jueves
Horario	: 10:00 a 13:00 hrs
Semestre que se dicta	: Segundo
Sala	: Computación Z06 - Subterráneo Edificio N°210

II. BREVE DESCRIPCIÓN DEL CURSO

Este curso está enfocado en entregar los principios de investigación reproducible en R, con énfasis en la recopilación y/o lectura de datos de forma reproducible y automatizada. Para esto se trabajará con bases de datos complejos, las cuales deberán ser transformadas y organizadas para optimizar su análisis. Se generarán documentos reproducibles integrando en un documento código, bibliografía, exploración y análisis de datos. Se culminará el curso con la generación de un manuscrito, presentación y/o documento interactivo reproducible.



III. OBJETIVOS

1. Conocer y entender el concepto de Investigación reproducible como un forma y filosofía de investigación que permite que las investigaciones sean más ordenadas y replicables, desde la toma de datos hasta la escritura de resultados.
2. Conocer y aplicar el concepto de pipeline, el cual permite generar una modularidad desde la toma de datos hasta la escritura de resultados, donde la corrección independiente de un paso tiene un efecto cascada sobre el resultado final.
3. Aprender buenas prácticas de recolección y estandarización de bases de datos, con la finalidad de optimizar el análisis de datos y la revisión de estas por pares.
4. Realizar análisis críticos de la naturaleza de los datos al realizar análisis exploratorios, que permitirán determinar la mejor forma de comprobar hipótesis asociadas a estas bases de datos.

Los principios de este curso están explicados en los siguientes libros gratuitos:

- Gandrud, Christopher. Reproducible Research with R and R Studio. CRC Press, 2013. Available for free in the following URL (<https://englianhui.files.wordpress.com/2016/01/reproducible-research-with-r-and-studio-2nd-edition.pdf>)
- Stodden, Victoria, Friedrich Leisch, and Roger D. Peng, eds. Implementing reproducible research. CRC Press, 2014. Available for free in the following URL (<http://web.stanford.edu/~vcs/papers/ijclp-STODDEN-2009.pdf>)

IV. CONTENIDOS

1. Investigación reproducible. Generación de documentos Rmarkdown en R, códigos en chunk y en texto. Generación de tablas en Rmd
2. Tidy Data, optimización de base de datos, limpieza y transformación de bases de datos.
3. El Hadleyverso y el tidyverso, el concepto de pipeline. Limpieza de datos complejos
4. Visualización de datos, visualizar datos vs. visualizar modelos. Insertar gráficos con leyenda en un documento Rmd
5. Creación de funciones propias y loops. Generación de funciones propias en R y loops
6. Escritura de manuscritos en R, transformación de documentos Rmd en un manuscrito
7. Presentaciones en R y generar documentos interactivos. Transformación de datos en una presentación o en una Shiny app. Realizar una presentación o aplicación en R.

V. METODOLOGÍA

Todas las clases serán prácticas y estarán divididas en dos partes: I. Clases de principios y herramientas, donde se presentarán los principios de investigación reproducible y tidy data,



junto con las herramientas actuales más utilizadas, y II. Clases aplicadas donde se trabajará con datos propios para desarrollar un documento reproducible. A los estudiantes que no cuenten con datos propios, se les será proporcionado un set de datos o se simularán dependiendo del caso.

Además se deberán generar informes y presentaciones siguiendo los principios de investigación reproducible, usando datos propios o entregados. Se realizará un informe final, en el cual se espera un trabajo.

Requerimientos de Software

La última versión de RStudio y R, también se requiere de la instalación de miktex (windows) o mactex (mac), por último, se necesita tener instalado git bash o equivalente, para windows revisar (<https://git-for-windows.github.io/>).

VI. EVALUACIÓN

1. Evaluación 1: Informe exploratorio de base de datos (20 puntos)
2. Evaluación 2: Presentación (20 puntos)
3. Evaluación 3: Informe final (60 puntos)

VII. BIBLIOGRAFÍA

Mínima

- Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten simple rules for reproducible computational research. *PLoS computational biology*, 9(10), e1003285.
- Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, 59(10), 1-23.

Complementaria

- Gandrud, C. (2013). *Reproducible research with R and R studio*. CRC Press.
- (<https://englianhu.files.wordpress.com/2016/01/reproducible-research-with-r-and-studio-2nd-edition.pdf>)
- Stodden, V., Guo, P., & Ma, Z. (2013). *Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals*.
- Stodden, V., Leisch, F., & Peng, R. D. (Eds.). (2014). *Implementing reproducible research*. CRC Press. (<http://web.stanford.edu/~vcs/papers/ijclp-STODDEN-2009.pdf>)



VIII. CALENDARIO DE ACTIVIDADES

CLASE	TEMA	LECTURA	PAQUETES	ACTIVIDAD
1	Investigación reproducible	Peng, Roger. "Reproducibility and research in computational science." <i>Science</i> 334.6060 (2011): 1226-1227.	swirl tidyverse	Generar un documento Rmarkdown en R
2	Tidy Data	Tidy data by Hadley Wickham, <i>Journal of Statistical Software</i> August 2014, Volume 59, Issue 10.	tidyverse lubridate	Hacer que tus datos sean limpios
3	El Hadleyverso y el tidyverso		tidyverse, broom, lubridate	El flujo del tidyverso. Limpiar datos complejos
4	Visualización de Datos	Wickham, H. (2010). A layered grammar of graphics. <i>Journal of Computational and Graphical Statistics</i> , 19(1), 3-28.	tidyverse	Visualizar tus datos vs visualizar tu modelo
5	Creación de funciones propias y loops			Generar funciones propias en R y loops
6	Escribir manuscritos en R	https://github.com/rstudio/rmarkdown	Rmd	Transformar documento Rmd en un manuscrito
7	Presentaciones en R y documentos interactivos		Shiny	Transforma tus datos en una presentación o en una Shiny app
8	Presentaciones			